Scenario Generation and Sampling Methods

Güzin Bayraksan
Tito Homem-de-Mello

SVAN 2016 – IMPA
May 9th, 2016

Asymptotic properties

Recap: we were studying what happens when we approximate the problem

$$\min_{x \in X} \{g(x) := \mathbb{E}[G(x,\xi)]\} \tag{SP}$$

by

$$\min_{x \in X} \left\{ \hat{g}_N(x) \;:=\; \frac{1}{N}\sum_{j=1}^{N} G(x,\xi^j) \right\}. \tag{SP$_N$}$$

Asymptotic properties

Let

$$\begin{aligned}\hat{x}_N &:= \text{an optimal solution of } (\mathrm{SP}_N)\\ S_N &:= \text{the set of optimal solutions of } (\mathrm{SP}_N)\\ \nu_N &:= \text{the optimal value of } (\mathrm{SP}_N)\end{aligned}$$

and

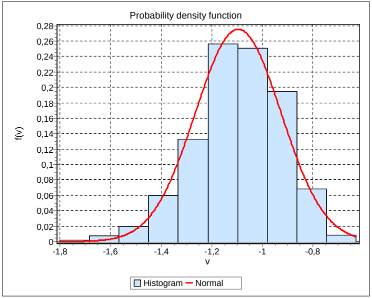
$$\begin{aligned}x^* &:= \text{an optimal solution of } (\mathrm{SP})\\ S^* &:= \text{the set of optimal solutions of } (\mathrm{SP})\\ \nu^* &:= \text{the optimal value of } (\mathrm{SP})\end{aligned}$$

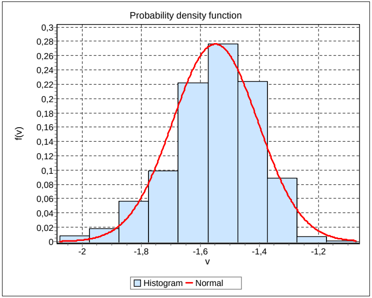
As the sample size N goes to infinity, does

- $\hat{x}_N$ converge to some x^*? How fast?
- S_N converge to the set S^*? How fast?
- ν_N converge to ν^*? How fast?

Rates of convergence

Let us take a look at the optimal values of the multiple runs shown in the graph we saw (exponential demand is on the left, uniform on the right)





The optimal values appear to be normally distributed across runs!

OK...but why does that help?

Rates of convergence

Again, how general is that behavior? Consider again a generic $G(x, \xi)$.

Theorem

Under assumptions of compactness of X and Lipschitz continuity of the function $G(\cdot, \xi)$ and assuming that the optimal solution x^* is unique, we have that

$$\sqrt{N}(\nu_N - \nu^*) \xrightarrow{d} \text{Normal}(0, \sigma(x^*)),$$

where $\sigma^2(x) = Var(G(x, \xi))$.

The symbol $\xrightarrow{d}$ means convergence in distribution.

Rates of convergence

Comments:

- The theorem says that, for N sufficiently large, ν_N is approximately normally distributed with mean ν^* and variance $\sigma^2(x^*)/N$.
- Of course, we don't know the values of ν^* and $\sigma^2(x^*)$; but the result is useful in saying that the optimal values converge at a rate of $1/\sqrt{N}$.
- When the optimal solution x^* is not unique, we replace the Normal distribution on the right hand side with

$$\inf_{x \in S^*} \text{Normal}(0, \sigma(x))$$

which is *no longer normally distributed*!

The bias problem

The theorem also leads to some important conclusions about the bias $\nu^* - \mathbb{E}[\nu_N]$.

What is this?

- Consider for the moment the case where $N = 1$. Then, we have

$$\mathbb{E}[\nu_1] \;=\; \mathbb{E}\left[\min_{x \in X} G(x,\xi)\right] \;\le\; \min_{x \in X} \mathbb{E}[G(x,\xi)] \;=\; \min_{x \in X} g(x) \;=\; \nu^*,$$

- It is easy to generalize the above inequality for arbitrary N, from which we conclude that

$$\mathbb{E}[\nu_N] \;\le\; \nu^*.$$

- That is, on the average, the approximating problem yields an optimal value that is below or at most equal to ν^*.

The bias problem (cont.)

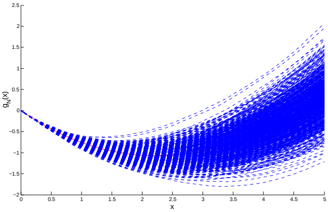
- In statistical terms, ν_N is a biased estimator of ν^* for all N.
- It is possible to show, however, that the bias $\nu^* - \mathbb{E}[\nu_N]$ decreases monotonically in N and goes to zero as N goes to infinity.
- For example, Freiman et al. (2010) compute the exact bias for the newsvendor problem when demand has uniform distribution on (0,1) as

$$\mathbb{E}[\nu_N] - \nu^* \;=\; \frac{\gamma(1-\gamma)}{2(N+1)},$$

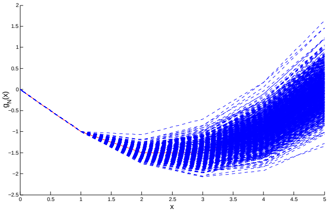
  so we see that in this case the bias is of order $1/N$.
- In general, the rate of convergence of bias for a stochastic program can take a variety of values N^{-p} for $p \in [1/2, \infty)$

Convergence result (cont.)

Recall the experiment where we solved (SP$_N$) (only for $N = 270$) multiple times, each time with a different sample path:



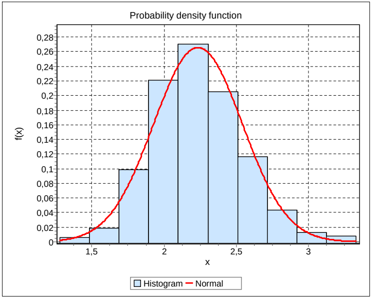
(a) Exponential demand

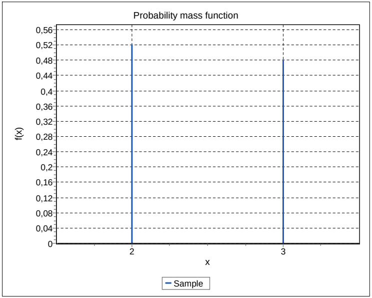


(b) Discrete uniform demand

Rates of convergence of optimal solutions

Let us take a look at the optimal *solutions* of the multiple runs shown in the graph (exponential demand is on the left, uniform on the right):





What do we observe?

Rates of convergence of optimal solutions

In the continuous case we see that the estimator $\hat{x}_N$ takes on many values but it concentrates around the true optimal solution $x^* = 2.23$, i.e., $P(\|\hat{x}_N - x^*\| > \varepsilon)$ goes to zero very fast.

This can be made precise:

Theorem

Under assumptions of compactness of X and certain boundedness and strong convexity assumptions on the function $G(\cdot, \xi)$, for any $\varepsilon > 0$ there exists a constant $\beta(\varepsilon) > 0$ such that

$$\limsup_{N\to\infty} \frac{1}{N} \log P(\|\hat{x}_N - x^*\| > \varepsilon) \;\leq\; -\beta(\varepsilon).$$

Essentially, this is saying that there exist constants C, $\beta(\varepsilon) > 0$ such that, asymptotically (and under appropriate conditions),

$$P(\|\hat{x}_N - x^*\| > \varepsilon) \;\leq\; Ce^{-\beta(\varepsilon)N}.$$

A glimpse on Large Deviations theory

Let $Z^1, Z^2, \ldots,$ be an i.i.d. sequence of random variables with common distribution P.

Define $\zeta_N := N^{-1}\sum_{j=1}^{N} Z^j$. Then for any real numbers a and $t > 0$ it holds that $P(\zeta_N \geq a) = P(e^{t\zeta_N} \geq e^{ta})$, and hence it follows from Chebyshev's inequality that

$$P(\zeta_N \geq a) \;\leq\; e^{-ta}\mathbb{E}\left[e^{t\zeta_N}\right] \;=\; e^{-ta}[M(t/N)]^N,$$

where $M(t) = \mathbb{E}[e^{tZ}]$ is the *moment generating function* of the distribution P.

A glimpse on Large Deviations theory

By taking the logarithm of both sides of the above inequality, changing variables $t' := t/N$, and minimizing over $t' \geq 0$, it follows that for $a \geq \mu := \mathbb{E}[Z]$ we have

$$\frac{1}{N} \log \left[P(\zeta_N \geq a) \right] \ \leq \ -I(a), \tag{1}$$

where

$$I(y) := \sup_{t \in \mathbb{R}} \{ty - \Lambda(t)\},$$

is the conjugate function of the logarithmic moment-generating function $\Lambda(t) := \log M(t)$.

The function I is called the (LD) rate function of Z. It is possible to show that both functions $\Lambda(\cdot)$ and $I(\cdot)$ are convex. Moreover, $I(\cdot)$ attains its minimum at $z = \mu$, and $I(\mu) = 0$.

A glimpse on Large Deviations theory

The constant $I(a)$ gives, in a sense, the best possible exponential rate at which the probability $P(\zeta_N \geq a)$ converges to zero for $a > \mu$:

Theorem (Cramér's Theorem)

If the moment generating function $M(t)$ is finite for all t, then

$$\lim_{N\to\infty} \frac{1}{N} \log\left[P\left(\zeta_N \geq a\right)\right] \;=\; -\inf_{y\geq a} I(y).$$

Large Deviations: example

Suppose $Z \sim \text{Binomial}(N, p)$. Want to compute $P(Z \geq N/2)$ for $p = 1/3$.

$$\lim_{N\to\infty} \frac{1}{N} \log\left[P\left(Z \geq N/2\right)\right] \;=\; -\inf_{y\geq 1/2} I_{\text{Bernoulli}}(y) \;=\; -I_{\text{Bernoulli}}(0.5)$$

where

$$I_{\text{Bernoulli}}(y) \;=\; y\log\left[\frac{(1-p)y}{p(1-y)}\right] - \log\left[\frac{(1-p)y}{(1-y)} + 1 - p\right]$$

and so

$$-I_{\text{Bernoulli}}(0.5) \;=\; -\log\left[\frac{(p^{-1}-1)^{1/2}}{2(1-p)}\right].$$

Thus when $p = 1/3$, we have for N large enough

$$P\left(Z \geq N/2\right) \;\approx\; e^{-0.0589N}.$$

The finite case

Consider now the special situation where the feasibility set X is finite, and suppose that the optimal solution x^* is *unique*, i.e., there exists $\varepsilon > 0$ such that $g(x^*) < g(x) - \varepsilon$ for all $x \in X \setminus \{x^*\}$

- Then the solution $\hat{x}_N$ of the approximating problem will coincide with x^* as long as

$$\hat{g}_N(x^*) < \hat{g}_N(x) \quad \text{for all } x \in X \setminus \{x^*\}.$$

- But, by LD theory, we know that, for any $x \in X$ and any $\delta > 0$, there exist constants C_x and $\beta_x > 0$ such that

$$P\left(|\hat{g}_N(x) - g(x)| > \delta\right) \;\approx\; C_x e^{-\beta_x}.$$

- Therefore in this case we can say that $P(\hat{x}_N \neq x^*)$ goes to zero exponentially fast.

The finite case

The result can be generalized for the case of multiple optimal solutions:

Theorem

Suppose that X is finite, and that for any $x \in X$ the moment generating function of $G(x, \cdot)$ is finite (in a neighborhood of zero). Then, there exists a constant $\beta > 0$ such that

$$P\left(S_N \not\subset S^*\right) \;\leq\; |X| e^{-\beta N}.$$

The discrete distribution case

Another special situation occurs in the following setting:

1. The feasibility set X is a convex closed polyhedron;
2. The distribution of ξ has finite support;
3. For every realization of ξ, the function $G(\cdot, \xi)$ is convex and piecewise linear.

One important class of problems that fits the above setting is that of *two-stage stochastic linear programs*.

The discrete distribution case

Under the three assumptions above, the problem behaves somewhat similarly to the case of finite X. The following result can be derived:

Theorem

Suppose that assumptions 1-3 above hold. Let $\mathcal{E}_N$ denote the event

$$\mathcal{E}_N \; := \; \{\textit{The set } S_N \textit{ is nonempty and forms a face of the set } S^*\}.$$

Then, there exists a constant $\beta > 0$ such that

$$\limsup_{N\to\infty} \frac{1}{N} \log P\left(\mathcal{E}_N^c\right) \; \leq \; -\beta.$$

Now the reason for behavior of $\hat{x}_N$ in the newsvendor problem with discrete uniform distribution is clear!